BMJ Oncology

# Bedside implications of the use of surrogate endpoints in solid and haematological cancers: implications for our reliance on PFS, DFS, ORR, MRD and more

Timothée Olivier,[1] Alyson Haslam ,[2] Dagney Ochoa,[2] Eduardo Fernandez ,[3] Vinay Prasad[2]

**To cite:** Olivier T, Haslam A, Ochoa D, *et al*. Bedside implications of the use of surrogate endpoints in solid and haematological cancers: implications for our reliance on PFS, DFS, ORR, MRD and more. *BMJ Oncology* 2024;**3**:e000364. doi:10.1136/bmjonc-2024-000364

TO and AH contributed equally.

Received 10 February 2024
Accepted 12 September 2024



[1]Department of Oncology, Geneva University Hospitals, Geneve, Switzerland
[2]Department of Epidemiology and Biostatistics, University of California San Francisco, San Francisco, California, USA
[3]Jane Anne Nohl Division of Hematology and Center for the Study of Blood Diseases, USC Norris Comprehensive Cancer Center, Los Angeles, California, USA

**Correspondence to**
Vinay Prasad;
vinayak.prasad@ucsf.edu

## ABSTRACT

Clinical endpoints, such as overall survival, directly measure relevant outcomes. Surrogate endpoints, in contrast, are intermediate, stand-in measures of various tumour-related metrics and include tumour growth, tumour shrinkage, blood results, etc. Surrogates may be a time point measurement, that is, tumour shrinkage at some point (eg, response rate) or biomarker-assessed disease status, measured at given time points (eg, circulating tumour DNA, ctDNA). They can also be measured over time, as with progression-free survival, which is the time until a patient presents with either disease progression or death. Surrogates are increasingly used in trials supporting the marketing authorisation of novel oncology drugs. Yet, the trial-level correlation between surrogates and clinical endpoints—meaning to which extent an improvement in the surrogate predicts an improvement in the direct endpoint—is often moderate to low. Here, we provide a comprehensive classification of surrogate endpoints: time point measurements and time-to-event endpoints in solid and haematological malignancies. Also, we discuss an overlooked aspect of the use of surrogates: the limitations of surrogates outside trial settings, at the bedside. Surrogates can result in the inappropriate stopping or switching of therapy. Surrogates can be used to usher in new strategies (eg, ctDNA in adjuvant treatment of colon cancer), which may erode patient outcomes. In liquid malignancies, surrogates can mislead us to use novel drugs and replace proven standards of care with costly medications. Surrogates can lead one to intensify treatment without clear improvement and possibly worsening quality of life. Clinicians should be aware of the role of surrogates in the development and regulation of drugs and how their use can carry real-world, bedside implications.

## INTRODUCTION

Clinical endpoints—such as overall survival and health-related quality of life—measure what is inherently meaningful to patients. In contrast, surrogate endpoints are stand-ins, measuring tumour growth rates or depth of tumour shrinkage, based on scans, serum protein assays, bone marrow biopsies, nucleotide sequencing, and other tools. While not specific to the monitoring and management of cancer, surrogates have been considered in tumour detection for cancer screening.[1] Ideally, these endpoints predict subsequent improvement in clinical outcomes or can help guide optimal care. To validate a surrogate endpoint, validation studies should reliably demonstrate the surrogate's ability to predict changes in meaningful outcomes in trial-level correlation studies, using multiple high-quality randomised studies reporting both overall survival and the surrogate outcome.[2]

For multiple reasons, surrogates are increasingly used in clinical trials, including those supporting marketing authorisations (ie, approvals) of novel drugs: some consider them a direct measure of antitumour effects, they may allow patients more rapid access to innovation, they may help in the interpretation of other cancer outcomes, and they may serve as useful heuristics in clinical practice to decide whether to modify treatment.[3] Yet, the trial-level correlation between surrogates and clinical endpoints—meaning to which extent an improvement in the surrogate translates or predicts an improvement in the direct endpoint—is often moderate to low.[4–6]

A key question that is often overlooked is how the use of surrogate endpoints in trials affects daily clinical practice. Here, we describe surrogate endpoints in solid and liquid (haematological) tumour settings, with measures of tumour shrinkage (response rate (RR) and minimal residual disease (MRD)) and time-to-event endpoints (progression-free survival (PFS) or relapse-free survival (RFS)). We describe the impact and potential

Table 1 Examples of surrogates, their definitions and strengths and limitations when applied at the bedside

| Endpoints | Definitions | Strengths | Limitations to be aware of in daily practice |
|---|---|---|---|
| Circulating tumour DNA (ctDNA) | ctDNA may be released by tumour cells into the bloodstream | ► A simple blood test<br>► Can be repeated over time | ► When implemented 'in addition' to standard practice, may unintentionally lead to more treatment without improving outcomes |
| Response rate | Response rate is a time point measurement reflecting the percentage of patients whose cancer has shrunk to a prespecified threshold (at least 30% two-dimensional tumour shrinkage according to RECIST) | ► Measurement of activity (the ability to shrink tumour)<br>► Operationally useful | ► 'Confirmation' of response is not always applied like in RECIST<br>► Unblinded assessment tends to overestimate response rates. Does not capture when the response occurred<br>► Significant response rates may never translate into clinical efficacy |
| Progression-free survival (PFS) | PFS is a time-to-event endpoint assessing the time from treatment initiation or randomisation to disease progression or death | ► May be considered a direct measure of antitumour effects<br>► May allow patients more rapid access to innovation<br>► May serve as useful heuristics in clinical practice to decide whether to modify treatment | ► Informative censoring should be explored, particularly when one arm is more toxic<br>► Switching therapy beyond progression may not always improve clinical outcomes<br>► Limited PFS gains may be even lower than those reported, because the event can occur at any time point between 2 CT scans<br>► PFS remains a surrogate, which has, in and of itself, little clinical relevance |
| Adjuvant endpoints | Time-to-event composite endpoints (DFS, RFS, EFS) | ► Grouping several events have the advantage of increasing statistical power, and allowing for smaller sample sizes | ► They may confuse results' interpretation, particularly when a detailed breakdown is not reported<br>► The clinician should look up the precise definition of the surrogate being used, as it may change between trials |

DFS, disease-free survival; EFS, event-free survival; RECIST, Response Evaluation Criteria in Solid Tumors; RFS, relapse-free survival.

limitations of the use of surrogates on treatment decisions occurring at the bedside (table 1).

### How to classify surrogate endpoints: time point measurements and time-to-event endpoints

Surrogates can be classified into two broad categories: time point measurements of antitumoural activity (eg, tumour shrinkage measurements or biomarker-assed disease status, measured at interval time points) and time-to-event endpoints. An additional layer of classification distinguishes between solid and haematological malignancies (liquid tumours). Solid tumours are organised by tissue type, depending on the origin of the cancer. Conversely, in haematological cancers, tumour cells are localised in the blood, the bone marrow or lymph nodes. Different surrogates are used depending on tumour extension, that is, whether the tumour is confined to the primary affected organ or has spread to distant parts of the body.[7] We illustrate this classification in table 2.

Metrics of antitumour activity—tumour shrinkage—constitute a first category of surrogates. In solid tumours, the most commonly used is RR. Other surrogates may be used, such as pathological complete response (pCR) in the neoadjuvant setting or ctDNA (circulating tumour DNA). In liquid tumours, the response is captured by MRD or other measures of abnormal cells in the blood. For haematological measures, a response based on tumour shrinkage is not captured.

Time-to-event endpoints, the second category of surrogates, are the time elapsed until the event is measured. They usually measure several metrics, including tumour response and survival, which are grouped into one single 'composite' endpoint. Time-to-event metrics include PFS, disease-free survival (DFS), RFS, metastasis-free survival and event-free survival (EFS),[8] among others. While we cannot discuss every surrogate marker, we discuss several of the more commonly used endpoints.

Table 2 Surrogates endpoints according to the type of tumour (solid or liquid) and the type of measurement (tumour shrinkage or time to event)

| Tumour types | | Surrogates measures | | Example |
|---|---|---|---|---|
| | | Tumour shrinkage | Time-to-event endpoints | Drug/tumour/trial name |
| Solid | Local | pCR (pathological complete response) | | Pertuzumab/breast/ NeoSphere |
| | | ctDNA | | Fluoropyrimidine or oxaliplatin-based chemotherapy/ colon cancer/DYNAMIC trial |
| | | | Disease-free survival | Osimertinib/non-small cell lung cancer/ADAURA trial |
| | Advanced | RR (response rate) | | Rociletinib/non-small cell lung cancer/ NCT01526928 |
| | | | Progression-free survival | Everolimus/breast/Bolero |
| Liquid/haematological | | CR (Complete Remission) | | Acute myeloid leukemia |
| | | MMR (major molecular response) | | Chronic myeloid leukaemia |
| | | MRD (minimal residual disease) | | VMP-Rd/multiple myeloma/ GEM2017FIT |

ctDNA, circulating tumour DNA.

## The use of surrogates in trials and the evidence for their use in clinical practice?

### Solid tumours (local)

#### *Pathological complete response*

In the neoadjuvant setting of localised cancers, systemic treatments are given before surgery to enable less invasive procedures and improve long-term outcomes.[9] For example, in breast cancer, neoadjuvant therapy can allow for a lumpectomy instead of a mastectomy, offering significant benefits. pCR, the absence of cancer cells on histopathology post-treatment,[10] is a positive prognostic marker, indicating better outcomes for those achieving pCR.[11] However, there is a distinction between prognostic biomarkers and surrogacy. The validation of surrogacy is attained when the change in the surrogate endpoint under therapy has a high correlation with long-term outcomes such as overall survival.[2]

Pertuzumab, an anti-HER2 antibody, was granted accelerated approval in the neoadjuvant setting of patients with localised HER-2-positive breast cancer. The approval was supported by NeoSphere, a phase II randomised controlled trial (RCT), with pCR as the primary endpoint.[12] The study showed higher rates of pCR with pertuzumab, trastuzumab, and docetaxel (chemotherapy), as compared with trastuzumab plus docetaxel (46% vs 29%). At the time of approval, the surrogacy of pCR in this setting was debated including within the Food and Drug Administration (FDA) itself,[13] however, the product was approved, mainly based on the known favourable prognosis of pCR.

As a postmarketing requirement, the phase III APHINITY RCT[14] showed that the addition of pertuzumab to chemotherapy and trastuzumab, in the adjuvant setting, led to a better 3-year invasive DFS (iDFS) of 94.1%, as compared with 93.2% in the control group (p=0.045). Primary cardiac events—a composite endpoint of cardiac adverse events including cardiac deaths—were more frequent when pertuzumab was added (0.7%), as compared with the placebo group (0.3%). Note the large difference in results. A 17-point improvement in pCR is linked to a less than 1% improvement in 3-year iDFS. The absolute difference in time-to-event endpoints like iDFS is an interesting metric as it directly correlates with the number needed to treat (100/absolute difference). In this example, the 0.9 absolute difference would translate into 111 patients needing to be treated to avoid one iDFS event after 3 years, a metric that can be directly useful during shared decision-making with the patient.

Obviously, given the baseline iDFS rates, which are high in both groups, a similar absolute improvement as seen with pCR could not have been observed. This highlights that surrogate measures should be interpreted in their own context, and a large difference in an early surrogate outcome like pCR may be associated with minimal improvement in other surrogates like iDFS. Importantly, overall survival, a secondary endpoint of APHINITY, was not improved in the third interim analysis.[15] With this longer follow-up, the absolute iDFS difference at 8 years is now 2.7, still resulting in a significant NNT=37.

When it comes to real-world practice, there are significant differences between patients enrolled in trials and those treated outside studies. A study conducted over about 125 000 oncology patients showed that 38% of them would not be eligible for trials.[16] Sanoff *et al* showed that

the median survival of patients with hepatocellular carcinoma receiving sorafenib in real life was half of that of patients receiving a placebo in the SHARP trial, the RCT that led to sorafenib approval.[17 18] This is often referred to as the 'efficacy-effectiveness gap'. In other words, any benefit seen in trials may not fully translate when applied to people treated outside trials.[19]

Similar discrepancies are observed with toxicity. In real life, patients often experience higher rates of adverse events than reported in trials.[20] When implementing the APHINITY strategy, the magnitude of the pertuzumab-driven incremental cardiac toxicity may be significantly higher.

In other words, APHINITY showed a limited absolute benefit in a surrogate measure, with no overall survival benefit and an increase in potentially severe toxicities. Yet, the approval was converted into a regular approval by the FDA. When applied in real life, the benefit could be less and the strategy more toxic than observed in the trial. In light of these considerations and uncertain health-related quality of life, clinicians should carefully assess surrogate validity prior to adopting novel strategies. If doctors change practice based on pCR, we have no guarantee that patients will live longer or better, but only that they will receive more drugs with more toxicity, time toxicity, and cost.

## Circulating tumour DNA

ctDNA can be released by tumour cells into the bloodstream, and as such, has become increasingly popular, offering many potential advantages over traditional tissue biopsy, including being a simple blood test that can be repeated over time.[21]

In early-stage colon cancer, the presence of ctDNA after surgical removal of the tumour is an adverse prognosis factor, that is, patients with ctDNA positivity after surgery bear a higher risk of recurrence.[22] Whether ctDNA is not only a prognostic marker but also a predictive marker (ie, it can be used to select patients that would benefit from adjuvant chemotherapy) was investigated in the DYNAMIC trial.

In patients with stage II colon cancer, the DYNAMIC RCT trial compared two strategies to determine subsequent adjuvant therapy. One strategy was the current care, where adjuvant chemotherapy was decided based on traditional clinicopathological characteristics. The novel strategy was guided by the ctDNA results.[23] The trial was presented as positive, showing non-inferior 2-year RFS with the ctDNA strategy, with a lower percentage of patients receiving chemotherapy (15% in the ctDNA group and 28% in the control group). However, we previously showed that implementing ctDNA in real-world practice could lead to the opposite goal of DYNAMIC, with more patients receiving chemotherapy.[24]

In colon cancer, a peritoneal invasion is a T4 in the tumor, node, and metastasis (TNM) staging and is considered a high-risk feature. Using traditional criteria, patients with T4 are usually eligible for adjuvant chemotherapy.[25 26] However, a key finding in DYNAMIC was that patients in the ctDNA strategy group with T4 tumours and a negative ctDNA result—consequently not receiving chemotherapy—had a lower 3-year RFS (81.3%) compared with those with a positive ctDNA (regardless of the T status) receiving chemotherapy (86.4%). In other words, if one were to apply the ctDNA strategy and dismiss the T4 factor to decide on chemotherapy, this would result in inferior outcomes for patients. Naturally, this finding led doctors to express reluctance to dismiss the T4 characteristic from their decision-making.[27]

In clinical practice, we suspect that clinicians will use ctDNA in addition to conventional clinicopathological features, and not in lieu of clinical characteristics, as was tested in DYNAMIC. The result would be that patients would be treated if they had either clinical characteristics or a positive ctDNA (ie, two different and not always overlapping ways to reach the decision to treat) which could potentially lead to more patients being treated. Accordingly, rates of adjuvant chemotherapy would be higher than in the study, diminishing any potential benefit of ctDNA-guided therapy.

Another concern in DYNAMIC was the type of chemotherapy being used. Fluoropyrimidines are the backbone of adjuvant chemotherapy, and oxaliplatin may sometimes be added. However, long-term neurotoxicity is a common concern with oxaliplatin. In DYNAMIC, 350% more patients received oxaliplatin in the ctDNA group than in the control, with 9.5% and 2.7% of patients receiving oxaliplatin, respectively.[23] Thus, ctDNA may paradoxically lead to the use of more toxic chemotherapy.[24 28 29]

The case of colon cancer illustrates the discrepancy between trial conditions and real-life regarding the use of surrogates. Trials testing ctDNA instead of clinical characteristics need to ensure that real patient benefit is being optimised; in the real world, doctors will often use new technology in addition to available data. New studies better tailored for the bedside are needed.

## Disease-free survival

DFS is the time a patient survives after treatment without signs or symptoms of cancer. In EGFR-mutant advanced or metastatic non-small cell lung cancer (NSCLC), osimertinib, a third-generation EGFR inhibitor, is considered the best first-line treatment based on the survival gain observed in the FLAURA trial, comparing osimertinib to gefitinib or erlotinib.[30]

When a treatment is effective in advanced settings, it may later be tested in the adjuvant setting. In contrast to advanced settings, where all patients have cancer and antitumour treatments are given, the adjuvant setting offers treatment to patients after surgical removal of early-stage cancer and who are deemed 'cancer-free'. Some tumours will never recur, and others will later present a relapse. Therefore, the goal of additional therapy in the adjuvant setting is to improve overall outcomes by changing the trajectory of tumours that would have recurred, while limiting harms to patients who would not have had

tumour recurrence. The question is whether exposing all patients to the drug is better than waiting for the relapse, in a fraction of them, to initiate the therapy.

This was investigated in the ADAURA trial where patients with resected stage IB-IIIA EGFR-mutated NSCLC were randomised to either 3 years of osimertinib or placebo. The trial showed an improvement in DFS in the osimertinib arm. While it was debated whether the DFS benefit, by itself, was not enough to change practice,[31] ADAURA ultimately showed an OS gain with osimertinib.[32]

The debate about whether adjuvant osimertinib should be used was not entirely settled by the ADAURA trial, simply because the trial does not mirror clinical practice. In ADAURA, the staging was inferior to standard practice (no mandatory PET-CT or brain MRI). This likely led to including patients with micrometastatic disease. Enrolling metastatic patients would automatically favour the osimertinib group: receiving a placebo for metastatic disease is not acceptable.

More concerning, the 5-year overall survival gain—10%—has to be contextualised. 37% of control patients experienced a recurrence (excluding deaths) and never received osimertinib when the tumours recurred. It raised the question of whether the survival gain (10%) would have been maintained if most of these patients (37%) had had access to osimertinib. Those numbers are explained by the global enrolment of ADAURA, including in countries with little or no access to the best standard of care on recurrence.[33 34]

The ADAURA trial highlights the issue with many contemporary trials, where the control arm—'on trial' and after the trial ends—deviates from best clinical care. A recent systematic review found that, in recent FDA approvals in the neoadjuvant or adjuvant settings (from 2018 to 2023), the postrecurrence treatment was considered inappropriate in 75% of trials where the data were reported.[35] The applicability to current practice of such trials is, therefore, questionable.

### Adjuvant time-to-event endpoints: volatile definitions

RFS, like DFS, only captures the primary malignancy and does not include the occurrence of secondary primary cancers other than the first cancer. Conversely, second cancers, in addition to primary cancers, may be captured by EFS endpoints. Nevertheless, women with breast cancer were estimated to have a 17% increased risk of developing a non-breast secondary malignancy.[36] As such, different definitions of composite endpoints modify their clinical interpretation.

Across studies, those endpoints lack universal definitions. Efforts, such as the DATECAN initiative, have aimed to reduce definition discrepancies.[37] Yet, no policy can ensure that studies are adhering to such consensus. Inconsistent definitions can also lead to imprecise estimates in meta-analyses.[38]

How are these endpoints integrated into clinical reasoning? Within a single breast cancer trial, any of the following outcomes equally counted as an 'iDFS' event: (1) invasive ipsilateral tumour recurrence, (2) invasive contralateral breast cancer, (3) local, (4) regional invasive recurrence, (5) distant recurrence or (6) death from any cause.[39]

From a clinical perspective, among these outcomes, some are amenable to cure, while others are lethal or potentially lethal.[40 41] Grouping several events has the advantage of increasing statistical power and allows for smaller sample sizes. However, they may confuse results' interpretation, particularly when a detailed breakdown is not reported. The clinician should look up the precise definition of the surrogate being used, as it may change from trial to trial.

## Solid tumours (advanced)

### Response rate

RR is a time point measurement reflecting the percentage of patients whose cancer has shrunk to a prespecified threshold, regardless when this occurs over time.[42] Approvals based on RR have allowed effective drugs to come to market earlier.

However, one issue with RR as an outcome is that it is based on interval scans that do not capture the moment at which the response is achieved. The response may be seen on the next CT scan, while in other cases, it may take months to be captured. In other words, the rapidity of response is missing from RR measurements. Clinicians may sometimes decide between different options based on RR data, with the false idea that higher RRs automatically translate into faster responses, which may not always hold. Time-to-response is also an important variable to consider in practice.

An additional consideration is that not all drugs that generate response—even robust ones—ultimately demonstrate clinical benefit. In that regard, the regulatory saga of tositumomab is noteworthy. Tositumomab was a monoclonal antibody targeting CD20, conjugated with a radioactive iodine compound delivering local radiation to CD20-positive cells. The compound showed a 68% RR in 40 patients, leading to its FDA approval in 2003 in patients with 'CD20 positive, follicular, non-Hodgkin's lymphoma, with and without transformation, whose disease is refractory to rituximab'.[43] The indication was expanded to rituximab-naïve patients in 2004 after 60 patients demonstrated a 47% RR in another single-arm study.

The postmarket required RCT was unable to accrue. The manufacturers proposed, instead, that a first-line RCT in patients with follicular lymphoma—the SWOG trial—could serve as a confirmatory trial. Because the setting was different from initial indications, the FDA warned the company that the trial might not satisfy the postmarketing requirements.

Ultimately, the SWOG trial, comparing tositumomab and rituximab, in addition to chemotherapy, reported a worse survival with tositumomab—93% 2-year survival—as compared with patients receiving rituximab (97%).[44]

After 11 years on the market, tositumomab was pulled off the market by the company, due to claims of a decline in sales.[45] The key message at the bedside is that even highly active drugs lack data to guide doctors in treatment decisions. Despite a high RR, the practising clinician never knew when or how to give I131 tositumumab in a way that could improve clinical outcomes. This same concern is present for many other drugs approved on RR and that lack RCT data showing survival benefit.[42]

Clinicians should be aware of the limitations of RR measurements in their decision-making. Those are (1) in real-life, 'confirmation' of response is not always applied like in RECIST, (2) unblinded assessment, which is that of daily practice, tends to overestimate RR, (3) RRs do not capture the rapidity of responses and (4) significant RRs may never translate into clinical efficacy. For some of these reasons, and omission of patients who are not able to receive subsequent imaging, waterfall plots visually overestimate response rate by approximately 12%.[46]

### Progression-free survival

PFS is a time-to-event endpoint assessing the time from treatment initiation or randomisation to disease progression or death.[47] PFS is a composite endpoint that not only measures tumour progression, but also death, although most PFS events are due to the former. There is the assumption that keeping the tumour from growing will improve survival time for the patient, but this is not always the case. In the absence of improved survival or quality of life, the improvement in surrogate outcomes alone has little value to the patient.[48] Consequently, through discussions, patients and physicians should be realistic about the value of drugs and how the benefit of the drug is determined.

Approvals based on PFS have led to the earlier approval of effective therapies. For example, ibrutinib was initially approved on PFS for patients with previously untreated chronic lymphocytic leukaemia. The approval was based on the results of the E1912 trial, and the overall survival results, while high in both arms, later showed the superiority of survival in patients assigned to the ibrutinib arm.[49]

In another example, osimertinib—a third-generation TKI—was approved based on a PFS benefit over earlier generation TKIs. There is ongoing debate in the oncology community about whether PFS should be used as a regulatory endpoint, thus allowing drugs to be available earlier to patients. On the one hand, osimertinib, after being initially approved on PFS, was later found to provide an overall survival benefit, and one could contend that patients would have been deprived of a beneficial drug had the approval not occurred until overall survival results were available. On the other hand, patients in the control arm of FLAURA, on progression, had limited access to osimertinib, which limits the validity and interpretability of the survival benefit.

It is important to understand the nuances of PFS and the factors related to its measurement to better contextualise results based on this surrogate outcome. As an example, BOLERO-2 was a phase III RCT, in patients with metastatic hormone-sensitive breast cancer, assessing the efficacy of everolimus plus exemestane (a hormone treatment) over exemestane alone. The study found a PFS benefit but never showed a survival gain.[50] The PFS plots displayed significantly higher rates of early censoring in the combination therapy group. This suggests that 'informative censoring' could have driven the reported PFS gain.

When patients withdraw from a trial or stop the experimental therapy long enough before undergoing the CT scan, they are 'censored'. This happens simply because the event (ie, progression) is no longer captured. However, important assumptions should be met to ensure that PFS comparison between arms is fair. One is that censoring rates should not be influenced by the allocated arm. Another assumption is that censored patients are generally similar, in terms of characteristics and likelihood of presenting the event, to patients remaining on trial. When those assumptions are violated, the censoring is referred to as 'informative', indicating that the rates of censoring carry information pertaining to the treatment group.

The most reasonable explanation for higher rates of censoring in the everolimus-exemestane arm of BOLERO-2 was increased toxicity. Patients who dropped out were most likely to be frail and bear an increased risk of presenting with the event, compared with uncensored patients. Exploring this hypothesis, it was shown that if patients in the experimental therapy had presented with the event instead of being censored, the PFS benefit would no longer be present.[51]

RECIST uses a 20% sum of diameters threshold to recommend changes in therapy when tumours grow beyond this metric. However, tumour progression does not equal a lack of anti-tumour activity: a therapy might still slow the tumour growth, even though beyond the arbitrary boundary of 'progression'. It was shown that sunitinib was able to reduce tumour growth beyond progression based on RECIST criteria. This could explain TKI sensitivity after progression.[52]

In other words, whether patients benefit from switching from sunitinib to another TKI should not be inferred based on the sole fact that patients progressed under sunitinib but should be prospectively tested. In fact, remaining on the same therapy after progression may be better than switching to a novel therapy that may be less potent, allowing for faster tumour growth.

Another limitation with PFS estimates is related to the time between CT scans. The CodeBreaK 200 trial randomised patients with G12C-mutant NSCLC after platinum-based chemotherapy and immunotherapy to receive either sotorasib—a KRAS G12C inhibitor—or docetaxel.[53] The trial reported a 5-week PFS benefit, a gain inferior to the 6-week CT scan intervals in the trial. This led the FDA to conduct an 'interval censoring analysis', where they randomly assigned the events between two CT scans and found that the PFS estimates could be

as low as 5 days. The CodeBreaK 200 trial had many other limitations, such as a substandard control arm and lack of power to capture overall survival benefits or even decrement.[54] However, the lesson here is that when facing limited PFS improvements, clinicians should always consider that the real gain could even be shorter.

Last, the arbitrary nature of progression—the '20%' boundary—is somehow disconnected from the patient experience: does a patient, while feeling well when the cancer has grown 19%, suddenly feel worse at 21%? The POLO trial investigated whether the maintenance of olaparib improved PFS in patients with germline BRCA mutation and metastatic pancreatic cancer.[55] The design of POLO had several limitations, mainly with a subpar control arm (it is not standard practice to stop chemotherapy in a patient who is responding) and survival not being the primary endpoint, despite the poor prognosis.[56] Overall, POLO showed an improvement in PFS, but no survival gain and added toxicity—olaparib is more toxic than placebo. POLO trial failed to improve clinical endpoints, yet the drug was approved by the US FDA. Clinicians face the challenging task of carefully interpreting the clinical relevance of surrogate improvements when considering novel strategies that may be harmful to patients.

The takeaway messages regarding PFS estimates when applied to daily practice are (1) the possibility of informative censoring should be explored, particularly when one arm is more toxic, (2) switching therapy beyond progression may not always be beneficial and should be assessed prospectively, (3) limited PFS gains may be even lower, simply because the event can occur at any time point between two CT scans and (4) PFS remain a surrogate, which has, in and of itself, little clinical relevance, and drugs approved on PFS, or other surrogate endpoints, should also provide survival data for confirmation. Adding toxicity without an improvement in clinical endpoints should generally be avoided.

## Haematological tumours

### Response rate

In liquid tumours, measures of antitumoural activity are typically measured through the percentage of bone marrow cancer cell reduction, circulating tumour markers or proteins, peripheral circulating cancer cells reduction, lymph nodes shrinkage, liver and spleen modification, or some combination of these, depending on cancer.

While there are examples of drugs approved initially based on complete remission rates (eg, azacitidine (AZA)–venetoclax for untreated acute myeloid leukaemia[57](AML)) and then later demonstrating improved overall survival, many drugs approved initially on a surrogate endpoint have unknown survival benefit. Hence, there are considerations when response rate (RR) is used as a surrogate. An example of surrogate complexity in haematology is seen in AML. In this disease, several surrogates may slightly differ in their definitions: complete remission (CR), CR with incomplete count recovery (CRi), CR with incomplete platelet recovery (CRp), and others. When grouped, they can constitute 'composite' endpoints, based on multiple single measurements.

LACEWING was a phase III trial that sought to evaluate the efficacy of adding the FLT-3 inhibitor gilteritinib to AZA versus AZA alone, in FLT-3 positive AML in patients not eligible for intensive chemotherapy. In the trial, the combination led to a statistically significant improvement in the composite of CR, CRi, and CRp (referred to as CRc in the trial); yet failed to yield a significant improvement in its primary endpoint of overall survival.[57 58]

The LACEWING example shows that grouping surrogates into a 'composite' surrogate can achieve statistical significance, yet clinical outcomes like survival may not be improved.

### Measurable residual disease

MRD is a tumour shrinkage measure, indicating the number of cancer cells remaining in the body after cancer treatment.[59] Because it is readily measurable and can be ascertained soon after treatment starts, many see MRD as a potential endpoint for use in drug approval. In other words, drugs that increase MRD rates may be initially approved, assuming that they will later prolong survival. Though, it is unknown whether MRD is a good surrogate for survival. To investigate this, Munshi *et al* performed an analysis asking if MRD-negative patients do better than MRD-positive patients with multiple myeloma.[60] Unfortunately, this analysis did not address the question of surrogacy. The authors did not explore whether in trials, drugs that increase MRD negativity later improve survival.[61] Surrogacy requires a specific form of trial-level analysis, which is not always performed correctly.[62] At times, those pushing for novel surrogates to be used for regulatory decisions have financial conflicts with manufacturers who benefit from lax regulation.

The lack of predictive value of MRD is further evidenced by the fact that single-agent drugs or drug combinations that achieve higher rates of MRD do not always lead to improvement in survival. One such example is the ongoing phase III trial GEM2017FIT which compares bortezomib, melphalan, and dexamethasone, followed by lenalidomide and dexamethasone (VMP-Rd) to carfilzomib, lenalidomide, and dexamethasone, with or without daratumumab (KRd/D-KRd) in 'fit', newly diagnosed patients with multiple myeloma (65–80 years of age). In this trial, a higher proportion of patients achieved MRD negativity in the D-KRd arm, yet survival was only numerically lower at 18 cycles of treatment in the intervention group of patients compared with patients assigned to KRd (95% vs 90%).[63] This finding highlights that MRD fails to capture the adverse effects of treatment that

may overcome the therapeutic benefit of achieving a deep remission.

There could be a point at which combining more drugs to achieve a higher MRD leads to higher toxicity and thus worse outcomes. In other words: chasing improvement with the use of surrogates may ultimately be detrimental to patients.

### Major molecular response

Major molecular response (MMR) is a tumour shrinkage measure that has been used to detect treatment failure when MRD is not met. However, like many surrogate endpoints, these measures are repeatedly used without proper surrogate measure validation; a strong correlation between improvements in the surrogate and clinical endpoints. This is the case in chronic myeloid leukaemia (CML), where MRD has repeatedly been shown to have no impact on overall survival. For example, one study showed that failure to achieve MRD optimal levels at 12 or 18 months was not associated with the probability of 5-year overall survival (96.4% vs 93.4% at 12 months and 95.6% vs 94.5% at 18 months, for MMR response and no MMR response, respectively).[38 64]

A common course of treatment for CML is imatinib. However, second-generation TKIs (2G-TKIs) have gained popularity due to faster and deeper molecular response, yet are associated with 30 times the cost of imatinib and higher toxicity, including cardiovascular, pulmonary, pancreatic, and hepatic toxicities. Additionally, many phase III trials have failed to establish a correlation between overall survival and MMR rates in CML, and there have been no high-quality trial-level correlation analyses to measure surrogacy for this marker. 2G-TKIs are a prime example of treatments that are widely justified by improved surrogate values, yet these surrogates fail to establish strong correlations with overall survival and other clinical endpoints, and it is unclear if better clinical outcomes are achieved.

Recently, asciminib, a third generation TKI, claimed superiority over both imatinib and a provider's choice of imatinib, dasatinib, and nilotinib.[65] However, the endpoint used was not overall survival or quality of life, but major molecular response at 48 weeks. While some praised this study as practice changing, it suffers from a key limitation. If investigators think better rates of major molecular response is sufficient to change practice, then imatinib should not be the control arm, as dasatinib and nilotinib already outperform imatinib. If instead, investigators believe that survival, quality of life, cost and toxicity should be the basis of practice, then the endpoint of the asciminib study – major molecular response – is inadequate. Investigators wish to have it both ways: they want an endpoint of unclear clinical significance (MMR) but also to test the novel drug against medications (imatinib) that are known to be suboptimal regarding that endpoint. Curiously, comparisons of asciminib vs. dasatinib and nilotinib (without imatinib) were relegated to a secondary objective. Clinicians, should be aware that many industry sponsored trials contain such prestidigitations.

## Conclusion

Surrogate endpoints have gained massive popularity in oncology, with PFS surpassing overall survival as the most common primary endpoint used in clinical trials.[4 66] The US FDA has steadily expanded the number of surrogate endpoints used for regulatory decisions.[67] A 2020 analysis described 194 novel drug authorizations since 1992 that were based on surrogates deemed acceptable by the FDA.[67] Largely, surrogates have been described in the context of regulatory decision-making; however, here we discuss the limitations of surrogates, which have implications for clinical practice, as they may not have intrinsic value to patients when there is a lack of survival benefit.

The use of surrogate endpoints has resulted in adding drugs in the solid tumour adjuvant setting with unclear value. Using surrogate endpoints to measure the clinical benefit of drugs for patients with cancer can result in the inappropriate stopping or switching of therapy. Their use can be used to usher in new strategies—ctDNA in the adjuvant setting of colon cancer—which may erode patient outcomes. In liquid malignancies, surrogates can mislead us to favour next-in-class drugs and replace proven standards of care with costly medications. Moreover, their use can lead to intensified treatment without any clear improvement (and possibly worsening) in patient outcomes. Clinicians should be aware that considerations of the use of surrogate endpoints are not limited to clinical trials, drug development and drug regulation. Their use can also result in real-world, bedside implications.

## Patient and public involvement

We did not have direct patient or public involvement since this was a review of the peer-reviewed literature, but we considered patient perspective and how our messages should be integrated into patient care.

**X** Timothée Olivier @Timothee_MD

**Contributors** Conception: VKP and TO; Writing–first draft: DO and TO. All authors reviewed and revised the manuscript. All authors approved the final draft.

**Funding** This research was funded by Arnold Ventures.

**Competing interests** VKP receives research funding from Arnold Ventures through a grant made to UCSF, and royalties for books and writing from Johns Hopkins Press, MedPage and the Free Press. He declares consultancy roles with UnitedHealthcare and OptumRX; He hosts the podcasts, Plenary Session, VPZD, Sensible Medicine, writes the newsletters, Sensible Medicine, the Drug Development Letter and VKP's Observations and Thoughts and runs the YouTube channel Vinay Prasad MD MPH, which collectively earn revenue on the platforms: Patreon, YouTube and Substack. All other authors have no conflicts to disclose.

**Patient and public involvement** Patients and/or the public were not involved in the design, or conduct, or reporting, or dissemination plans of this research.

**Patient consent for publication** Not applicable.

**Ethics approval** In accordance with 45 CFR §46.102(f), we did not submit this study to institutional review board approval because it involved publicly available data and did not involve individual patient data.

**Provenance and peer review** Commissioned; externally peer reviewed.

**Data availability statement** No data are available.



**ORCID iDs**
Alyson Haslam http://orcid.org/0000-0002-7876-3978
Eduardo Fernandez http://orcid.org/0000-0002-5489-5095

## REFERENCES

1 Webb AB, Berg CD, Castle PE, *et al*. Considerations for using potential surrogate endpoints in cancer screening trials. *Lancet Oncol* 2024;25:e183–92.
2 Kemp R, Prasad V. Surrogate endpoints in oncology: when are they acceptable for regulatory and clinical decisions, and are they currently overused? *BMC Med* 2017;15.
3 Parast L, Tian L, Cai T, *et al*. Can earlier biomarker measurements explain a treatment effect on diabetes incidence? A robust comparison of five surrogate markers. *BMJ Open Diabetes Res Care* 2023;11:e003585.
4 Prasad V, Kim C, Burotto M, *et al*. The Strength of Association Between Surrogate End Points and Survival in Oncology: A Systematic Review of Trial-Level Meta-analyses. *JAMA Intern Med* 2015;175:1389–98.
5 Haslam A, Hey SP, Gill J, *et al*. A systematic review of trial-level meta-analyses measuring the strength of association between surrogate end-points and overall survival in oncology. *Eur J Cancer* 2019;106:196–211.
6 Hwang TJ, Gyawali B. Association between progression-free survival and patients' quality of life in cancer clinical trials. *Int J Cancer* 2019;144:1746–51.
7 Cooper GM. *The Development and Causes of Cancer*. 2nd edn. Sinauer Associates, 2000.
8 Walia A, Tuia J, Prasad V. Progression-free survival, disease-free survival and other composite end points in oncology: improved reporting is needed. *Nat Rev Clin Oncol* 2023;20:885–95.
9 Luo J, Prasad V. The US Food and Drug Administration's use of pathologic complete response as regulatory endpoint: Did it pay off? *J Cancer Policy* 2018;16:49–51.
10 Conforti F, Pala L, Sala I, *et al*. Evaluation of pathological complete response as surrogate endpoint in neoadjuvant randomised clinical trials of early stage breast cancer: systematic review and meta-analysis. *BMJ* 2021;375:e066381.
11 Ivanovic N, Bjelica D, Loboda B, *et al*. Changing the role of pCR in breast cancer treatment - an unjustifiable interpretation of a good prognostic factor as a 'factor for a good prognosis. *Front Oncol* 2023;13.
12 Gianni L, Pienkowski T, Im Y-H, *et al*. Efficacy and safety of neoadjuvant pertuzumab and trastuzumab in women with locally advanced, inflammatory, or early HER2-positive breast cancer (NeoSphere): a randomised multicentre, open-label, phase 2 trial. *Lancet Oncol* 2012;13:25–32.
13 Amiri-Kordestani L, Wedam S, Zhang L, *et al*. First FDA approval of neoadjuvant therapy for breast cancer: pertuzumab for the treatment of patients with HER2-positive breast cancer. *Clin Cancer Res* 2014;20:5359–64.
14 von Minckwitz G, Procter M, de Azambuja E, *et al*. Adjuvant Pertuzumab and Trastuzumab in Early HER2-Positive Breast Cancer. *N Engl J Med* 2017;377:122–31.
15 Loibl S, Jassem J, Sonnenblick A, *et al*. Adjuvant Pertuzumab and Trastuzumab in Early Human Epidermal Growth Factor Receptor 2-Positive Breast Cancer in the APHINITY Trial: Third Interim Overall Survival Analysis With Efficacy Update. *J Clin Oncol* 2024.:JCO2302505.
16 Karim S, Xu Y, Kong S, *et al*. Generalisability of Common Oncology Clinical Trial Eligibility Criteria in the Real World. *Clin Oncol (R Coll Radiol)* 2019;31:e160–6.
17 Sanoff HK, Chang Y, Lund JL, *et al*. Sorafenib Effectiveness in Advanced Hepatocellular Carcinoma. *Oncologist* 2016;21:1113–20.
18 Llovet JM, Ricci S, Mazzaferro V, *et al*. Sorafenib in advanced hepatocellular carcinoma. *N Engl J Med* 2008;359:378–90.
19 Mailankody S, Prasad V. Overall Survival in Cancer Drug Trials as a New Surrogate End Point for Overall Survival in the Real World. *JAMA Oncol* 2017;3:889–90.
20 Hanlon P, Corcoran N, Rughani G, *et al*. Observed and expected serious adverse event rates in randomised clinical trials for hypertension: an observational study comparing trials that do and do not focus on older people. *Lancet Healthy Longev* 2021;2:e398–406.
21 Cheng F, Su L, Qian C. Circulating tumor DNA: a promising biomarker in the liquid biopsy of cancer. *Oncotarget* 2016;7:48832–41.
22 Tie J, Wang Y, Tomasetti C, *et al*. Circulating tumor DNA analysis detects minimal residual disease and predicts recurrence in patients with stage II colon cancer. *Sci Transl Med* 2016;8.
23 Tie J, Cohen JD, Lahouel K, *et al*. Circulating Tumor DNA Analysis Guiding Adjuvant Therapy in Stage II Colon Cancer. *N Engl J Med* 2022;386:2261–72.
24 Olivier T, Haslam A, Prasad V. Additional considerations before using a ctDNA-guided approach for informing adjuvant chemotherapy in colorectal cancer. *BMC Med* 2023;21:344.
25 Argilés G, Tabernero J, Labianca R, *et al*. Localised colon cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Ann Oncol* 2020;31:1291–305.
26 Baxter NN, Kennedy EB, Bergsland E, *et al*. Adjuvant Therapy for Stage II Colon Cancer: ASCO Guideline Update. *J Clin Oncol* 2022;40:892–910.
27 Olivier T, Prasad V. Molecular testing to deliver personalized chemotherapy recommendations: risking over and undertreatment. *BMC Med* 2022;20:392.
28 Kim H, Park KU. Clinical Circulating Tumor DNA Testing for Precision Oncology. *Cancer Res Treat* 2023;55:351–66.
29 Gibbs P, Hong W, Tie J. Re: Molecular testing to deliver personalised chemotherapy recommendations. *BMC Med* 2023;21.
30 Soria J-C, Ohe Y, Vansteenkiste J, *et al*. Osimertinib in Untreated EGFR-Mutated Advanced Non-Small-Cell Lung Cancer. *N Engl J Med* 2018;378:113–25.
31 Addeo A, Banna GL, Friedlaender A. ADAURA: Mature Enough for Publication, Not for Prime Time. *Oncologist* 2021;26:266–8.
32 Tsuboi M, Herbst RS, John T, *et al*. Overall Survival with Osimertinib in Resected *EGFR*-Mutated NSCLC. *N Engl J Med* 2023;389:137–47.
33 West HJ, Pennell NA. Lessons from ADAURA: Can we improve on a positive trial? *Cancer* 2024;130:659–64.
34 Olivier T, Haslam A, Prasad V. Post-progression treatment in cancer randomized trials: a cross-sectional study of trials leading to FDA approval and published trials between 2018 and 2020. *BMC Cancer* 2023;23:448.
35 Olivier T, Haslam A, Prasad V. Postrecurrence Treatment in Neoadjuvant or Adjuvant FDA Registration Trials: A Systematic Review. *JAMA Oncol* 2024;10:1055–9.
36 Molina-Montes E, Requena M, Sánchez-Cantalejo E, *et al*. Risk of second cancers cancer after a first primary breast cancer: a systematic review and meta-analysis. *Gynecol Oncol* 2015;136:158–71.
37 Gourgou-Bourgade S, Cameron D, Poortmans P, *et al*. Guidelines for time-to-event end point definitions in breast cancer trials: results of the DATECAN initiative (Definition for the Assessment of Time-to-event Endpoints in CANcer trials). *Ann Oncol* 2015;26:873–9.
38 Walia A, Prasad V. Is it time to reconsider molecular response milestones in chronic myeloid leukemia? *Am J Hematol* 2023;98:562–3.
39 Tolaney SM, Garrett-Mayer E, White J, *et al*. Updated Standardized Definitions for Efficacy End Points (STEEP) in Adjuvant Breast Cancer Clinical Trials: STEEP Version 2.0. *J Clin Oncol* 2021;39:2720–31.
40 Haslam A, Ranganathan S, Prasad V, *et al*. CDK4/6 inhibitors as adjuvant therapy in early breast cancer? Uncertain benefits, guaranteed harms. *Eur J Cancer* 2024;207:114192.
41 Martin M, Holmes FA, Ejlertsen B, *et al*. Neratinib after trastuzumab-based adjuvant therapy in HER2-positive breast cancer (ExteNET): 5-year analysis of a randomised, double-blind, placebo-controlled, phase 3 trial. *Lancet Oncol* 2017;18:1688–700.
42 Chen EY, Raghunathan V, Prasad V. An Overview of Cancer Drugs Approved by the US Food and Drug Administration Based on the Surrogate End Point of Response Rate. *JAMA Intern Med* 2019;179:915–21.
43 Corixa Corporation. BEXXAR (tositumomab and iodine I 131 tositumomab. 2003.
44 Press OW, Unger JM, Rimsza LM, *et al*. Phase III randomized intergroup trial of CHOP plus rituximab compared with CHOP chemotherapy plus (131)iodine-tositumomab for previously untreated follicular non-Hodgkin lymphoma: SWOG S0016. *J Clin Oncol* 2013;31:314–20.

45 Prasad V. The withdrawal of drugs for commercial reasons: the incomplete story of tositumomab. *JAMA Intern Med* 2014;174:1887–8.
46 Kim MS, Prasad V. Assessment of Accuracy of Waterfall Plot Representations of Response Rates in Cancer Treatment Published in Medical Journals. *JAMA Netw Open* 2019;PMCID:e193981.
47 Schultz KR, Bowman WP, Aledo A, *et al*. Improved early event-free survival with imatinib in Philadelphia chromosome-positive acute lymphoblastic leukemia: a children's oncology group study. *J Clin Oncol* 2009;27:5175–81.
48 Robinson AG, O'Donnell J, Booth C, *et al*. Patient perspectives of value of delayed disease progression on imaging (imaging PFS). A treatment trade-off experiment. *J Cancer Policy* 2021;30:100301.
49 Shanafelt TD, Wang XV, Kay NE, *et al*. Ibrutinib–Rituximab or Chemoimmunotherapy for Chronic Lymphocytic Leukemia. *N Engl J Med* 2019;381:432–43.
50 Piccart M, Hortobagyi GN, Campone M, *et al*. Everolimus plus exemestane for hormone-receptor-positive, human epidermal growth factor receptor-2-negative advanced breast cancer: overall survival results from BOLERO-2†. *Ann Oncol* 2014;25:S0923-7534(19)36905-4:2357–62:.
51 Prasad V, Bilal U. The role of censoring on progression free survival: oncologist discretion advised. *Eur J Cancer* 2015;51:2269–71.
52 Burotto M, Wilkerson J, Stein W, *et al*. Continuing a cancer treatment despite tumor growth may be valuable: sunitinib in renal cell carcinoma as example. *PLoS One* 2014;9:e96316.
53 de Langen AJ, Johnson ML, Mazieres J, *et al*. Sotorasib versus docetaxel for previously treated non-small-cell lung cancer with KRASG12C mutation: a randomised, open-label, phase 3 trial. *The Lancet* 2023;401:733–46.
54 Ranganathan S, Prasad V, Olivier T. The fate of sotorasib: a regulatory failure potentially harming patients. *Lancet Oncol* 2024;25:549–52.
55 Golan T, Hammel P, Reni M, *et al*. Maintenance Olaparib for Germline *BRCA*-Mutated Metastatic Pancreatic Cancer. *N Engl J Med* 2019;381:317–27.
56 Nishikawa G, Booth C, Prasad V. Olaparib for BRCA mutant pancreas cancer: Should the POLO trial change clinical practice? *Cancer* 2020;126:4087–8.
57 DiNardo CD, Jonas BA, Pullarkat V, *et al*. Azacitidine and Venetoclax in Previously Untreated Acute Myeloid Leukemia. *N Engl J Med* 2020;383:617–29.
58 Wang ES, Montesinos P, Minden MD, *et al*. Phase 3 trial of gilteritinib plus azacitidine vs azacitidine for newly diagnosed FLT3mut+ AML ineligible for intensive chemotherapy. *Blood* 2022;140:1845–57.
59 MedGen - NCBI. Residual disease (concept id: c0543478). NCBI. Available: https://www.ncbi.nlm.nih.gov/medgen/108162#:~:text=Definition [Accessed 26 Jun 2024].
60 Munshi NC, Avet-Loiseau H, Rawstron AC, *et al*. Association of Minimal Residual Disease With Superior Survival Outcomes in Patients With Multiple Myeloma: A Meta-analysis. *JAMA Oncol* 2017;3:28–35.
61 Little RF, McShane LM, Freidlin B. Myeloma Minimal Residual Disease and Surrogacy. *JAMA Oncol* 2017;3:1136.
62 Munshi N, Samur MK, Gregory W. Myeloma Minimal Residual Disease and Surrogacy—Reply. *JAMA Oncol* 2017;3:1136.
63 Mateos M-V, Paiva B, Teresa Cedena Romer M, *et al*. GEM2017FIT trial: induction therapy with bortezomib-melphalan and prednisone (vmp) followed by lenalidomide and dexamethasone (rd) versus carfilzomib, lenalidomide and dexamethasone (krd) plus/minus daratumumab (d), 18 cycles, followed by consolidation and maintenance therapy with lenalidomide and daratumumab: phase iii, multicenter, randomized trial for elderly fit newly diagnosed multiple myeloma (ndmm) patients aged between 65 and 80 years. ASH; 2023. Available: https://ash.confex.com/ash/2023/webprogram/Paper179866.html [Accessed 26 Jun 2024].
64 Marin D, Milojkovic D, Olavarria E, *et al*. European LeukemiaNet criteria for failure or suboptimal response reliably identify patients with CML in early chronic phase treated with imatinib whose eventual outcome is poor. *Blood* 2008;112:4437–44.
65 Hochhaus A, Wang J, Kim DW, *et al*. ASC4FIRST Investigators. *Asciminib in Newly Diagnosed Chronic Myeloid Leukemia N Engl J Med* 2024;391:885–98.
66 Del Paggio JC, Berry JS, Hopman WM, *et al*. Evolution of the Randomized Clinical Trial in the Era of Precision Oncology. *JAMA Oncol* 2021;7:728–34.
67 Chen EY, Haslam A, Prasad V. FDA Acceptance of Surrogate End Points for Cancer Drug Approval: 1992-2019. *JAMA Intern Med* 2020;PMCID:912–4.